Document: Project Proposal
Author: ivm
Date: 11/26/2003
Modified: 1/9/2004

Proposal

Introduction

The purpose of this project is to build a hardware and software infrastructure for network monitoring and status information currently collected and maintained by Data Communications group and Computer Security Team. Such infrastructure will help organize the data collection, storage and access, increase the efficiency of inter-organizational communication and simplify further development of network monitoring and analysis tools.

Design Proposal

Information Flow

The following components of the network monitoring and state information will be maintained and made available through the Network Information and Management Infrastructure:

- Network transactions information. This information is collected from border routers and in the future will be collected from other routers on the LAN. Transaction information will be available in two forms:
    - From a relational database. The database will store detailed short-term history of network transactions and long-term history in some sort of compiled form. Client will have to periodically poll the database for updates at the appropriate frequency.
    - Through synchronous data feed. Clients that need semi-real time access to transaction data will be able to "subscribe" to receive UDP-based feed of data.
- Network device identification and location information. This is the information collected from such sources as VPN and DHCP server, routers and switches used to map IP address of a network device to MAC address, user or administrator identification, physical node identification and location. This information will be stored and maintained in the relational database.
- Network state and workflow information. This is current status and historical information about recent and future network maintenance and security related actions performed on individual network devices or network device groups. For example, in case of remediation of a vulnerability detected by a security scan, it is necessary to record and time-stamp information on the vulnerability found, current status of the node, whether it was blocked or not, some sort of record of the communication with the owner or administrator of the node, record of actions taken by data communications group and the security team, etc.

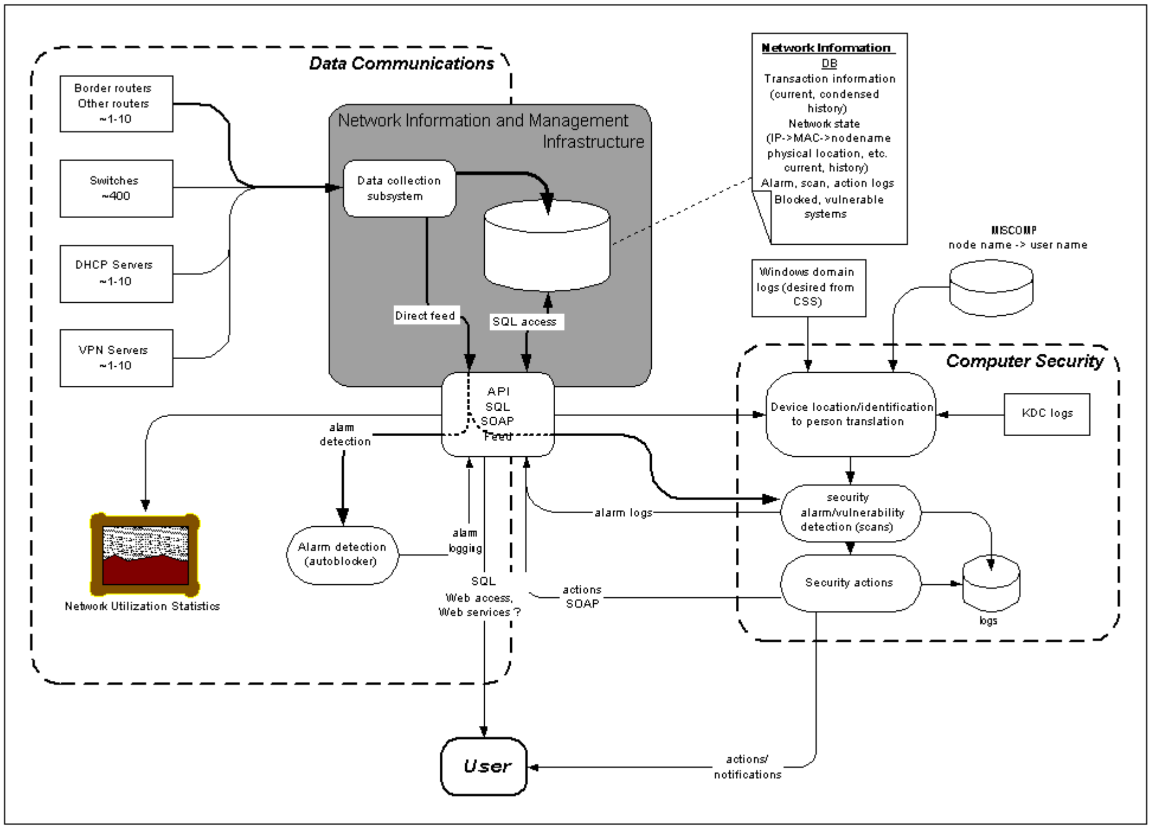
Depending on the nature of the data, some information will have to be protected from unauthorized access using some strong authentication mechanism. For example, currently network transaction information is available only to limited number of people from data communications and computer security teams. On the other hand, it seems to be necessary to give wider access to the workflow-related portion of the data, for example to speed-up the communication between network node administrator, computer security team and data communications.

Data Communications group will provide SOAP-based interface for Security Team to request changes in the network state (block or unblock access to certain node or nodes as needed for security reasons).

One of the first stages of the project will be analysis of the problem domain. The goal of this analysis step will be to isolate and more clearly identify elements of the data, their nature, lifetime and behavior, relationship between them. This information will be used to develop database schema and identify requirements to the interfaces needed to access and maintain the data.

Design

Proposed design is outlined in the following picture:


Data Communications
Border routers
Other routers
~1-10
Switches
~400
DHCP Servers
~1-10
VPN Servers
~1-10
Network Information and Management Infrastructure
Data collection subsystem
Direct feed
SQL access
API
SQL
SOAP
Feed
alarm detection
Alarm detection (autoblocker)
alarm logging
Network Utilization Statistics
SQL
Web access,
Web services ?
actions
SOAP
alarm logs
Network Information
DB
Transaction information (current, condensed history)
Network state (IP->MAC->nodename physical location, etc. current, history)
Alarm, scan, action logs
Blocked, vulnerable systems
MISCOMP
node name -> user name
Windows domain logs (desired from CSS)
Computer Security
Device location/identification to person translation
KDC logs
security alarm/vulnerability detection (scans)
Security actions
logs
User
actions/
notifications

The Infrastructure consists of 3 major parts:

- Data Collection Subsystem. This is set of tools used to collect information from various sources such as routers, switches, DHCP logs, etc. and store this information in the Database. It also can be used to collect information from other sources, for example to maintain secondary "local" copy or some sort of compilation of data stored in MISCOMP database.
- Network Information Database is a relational database where 3 described components of the network monitoring information will be stored. Several database solutions are available for consideration: Postgres, MySQL, Oracle.
- Interface will consist of the following parts:
    - Local and remote SQL access. This interface will have appropriate client authentication/authorization mechanism to protect data from unauthorized use.
    - Application Programmer's Interface. This is direct or network access provided in terms of problem-specific rather than relational database abstractions. API will be used primarily for building local applications.
    - Web Services/SOAP. This will be primary interface for simple requests from remote clients. Some authentication mechanism (e.g. Kerberos- or PKI-based) must be developed for this interface.
    - Direct transaction data feed. This portion of interface will be used by clients that need this information in real time. It has to be protected from unauthorized use.

Implementation

Hardware Implementation

The Network Information Database will have to reside on relatively big and fast computer. Bulk of the data is network transaction history. Currently this data is collected at the rate of ~10GB per day. This number is expected to grow in the future. If the data is to be stored for ~100 days before it is reduced for longer-term storage, then the computer must have ~1TB disk only for the network transaction data. One or two extra ~1TB disks will be used for short-term backups of the database. Also this computer will have to have fast network interface to allow:

- Access to the data
- Data backups over the network
- Data uploading

Key components of Data Collection Subsystem and some components of the Interface may run on the same computer as the database, but it may be more practical to place them on a separate computer or computers. These computers do not have to have large disk space capacity, but will have to have relatively fast network interface and may require fast CPU.

Software Implementation

- Data Collection Subsystem. Most of the components of it already exist. They will have to be modified to store data into the Database.
- Database. The structure of the data suggests that relational database schema is the most natural way of data representation in this case. Several software solutions are being considered. They are PostgresSQL, MySQL, Oracle. Oracle seems to be the most powerful and the most expensive solution. Free solutions like first two, if they can provide adequate performance, seem to be better choice. Additional research and testing in this area needs to be done.
- Interface. Implementation of local and network database interface will be determined by the choice of the database solution. SOAP/Web Services interface can be implemented in many languages such as C, C++, Python and Java. There are many packages available in public domain, which can be used to develop web-based interface. They are ZOPE, Tomcat, Matrix, Remedy, etc. just to name a few of those that come to mind. Some research in this area may be necessary to find the best middleware for our needs.